UDC: 004.8:174
DOI: 10.69899/limes-plus-en-24212-3099r
Original scientific work

ETHICAL IMPLICATIONS AND SOCIAL CHALLENGES OF ARTIFICIAL INTELLIGENCE DEVELOPMENT TOWARDS ARTIFICIAL GENERAL INTELLIGENCE

Viktor Radun[1]
Faculty of Contemporary Arts, Belgrade; University Business Academy, Novi Sad

Abstract: *This paper explains the ethical implications of artificial intelligence (AI) development towards achieving the level of artificial general intelligence (AGI) and analyzes the need for its social control. With the acceleration and intensification of AI growth and development, especially with the ongoing AI race, the transition from narrow AI to AGI becomes certain. The achievement of generative AI, which climaxes with chatbots (such as ChatGPT and others), transforms AI into a machine capable of creation. Although this AI application still appears relatively limited by algorithms, its learning ability is remarkable, and continuous advancements and the launch of increasingly sophisticated versions bring it ever closer to the AGI model. Each day brings us closer to that moment, which will signify AI's transition from narrow AI to AGI. Unlike narrow AI, AGI deeply delves into the realm of ethics, and interpersonal and social relationships. Regulating AI-related policy and legally controlling*

1 Imejl-adresa: viktor.radun@fsu.edu.rs

AI represents one of the most serious and complex issues. In recent years, the community, led by corporate executives developing AI, prominent experts, researchers, scientists, writers, and other stakeholders, has made significant steps towards raising public awareness of the risks posed by advanced AI and making decisions, initiatives, and measures for monitoring, analyzing, and socially controlling the use of AI.

Keywords: *Artificial Intelligence, Artificial General Intelligence, AGI, ethical implications, social control, risks.*

INTRODUCTION

Artificial Intelligence (AI) is a complex scientific and technological field, a branch of computer science that emerged from efforts to develop intelligent technology that would simulate the workings of the human brain and human intelligence. The definition of artificial intelligence has evolved over time in tandem with its development. According to John McCarthy (McCarthy, 2007), the creator of the term, artificial intelligence is "the science and engineering of making intelligent machines, particularly intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable".

According to Investopedia (Investopedia, 2024), "Artificial Intelligence technology enables computers and machines to mimic human intelligence and problem-solving tasks. The ideal characteristic of artificial intelligence is its ability to rationalize and take actions to achieve a specific goal".

McKinsey (McKinsey & Company, 2024) provides a broader definition, explaining AI as "the ability of a machine to perform cognitive functions associated with human minds, such as perception, reasoning, learning, interaction with the environment, problem-solving, and even creativity".

The aim of AI development is for technology to simulate human intelligence. AI is considered a general-purpose, universal technology, given its penetration and ease of efficient application across all sectors and areas of the

economy and society. It serves as a universal carrier and driver of technological development. AI is the leading technology among the new technologies that form the core of the Fourth Industrial Revolution (4IR). According to the Global Risks Report 2017 (World Economic Forum, 2017), twelve essential technologies make up the technological core of 4IR: a) 3D printing; b) advanced materials and nanomaterials; c) AI and robotics; d) biotechnology; e) energy capture, storage, and transmission; f) blockchain and distributed ledger; g) geoengineering; h) the Internet of Things; i) neurotechnology; j) new computing technologies; k) space technologies, and l) virtual and augmented technology.

Artificial intelligence is not a single technology but rather a set of specific technologies (sub-technologies) designed to perform specific tasks. In the AI spectrum, we find the following sub-technologies (Takyar, 2024):

- Machine Learning;
- Natural Language Processing – NLP;
- Computer Vision;
- Deep Learning;
- Generative Models;
- Expert Systems, and others.

DEVELOPMENT OF ARTIFICIAL INTELLIGENCE AND THE DEFINITION OF ARTIFICIAL GENERAL INTELLIGENCE

One of the primary characteristics of AI, like other 4IR technologies, is exponential growth. According to the most significant parameters, AI is experiencing exponential growth and development. This is true for the growth of computing power required to train generative AI models, which is of an exponential nature.

Figure 1 (Roser, 2022) shows the relationship between the growth of training computing power and the growth of AI systems' capabilities. The analysis of these two growths shows that in the pre-deep learning era, the growth of computing power for training AI systems followed Moore's Law, doubling approximately every 20 months. Since 2010, in the era of deep learning, this

growth has accelerated further, with computing power now doubling every six months.

Figure 1: Relationship between AI system training computation growth and system power over time.

The global AI market is substantial and experiencing rapid expansion. It is estimated that the total global AI market will be worth USD 538.13 billion in 2023, with expectations to reach USD 2,575.16 billion by 2032, marking a projected annual growth rate of 19% between 2023 and 2032.

According to generally accepted categorization, AI is divided into artificial general intelligence (AGI), artificial narrow intelligence (ANI), and artificial superintelligence (ASI). Instead of AGI (ANI) and ANI (ASI), some researchers use the terms “strong” and “weak” AI. This categorization is based on AI’s evolutionary potential, implying the continuous development of AI, which, at a critical point, can transition from the level of ANI, where it currently resides, to AGI, and subsequently, at the next critical point, to superintelligence and the singularity. Today, we encounter only narrow AI, whose applications and systems can perform certain specific intellectual tasks or exhibit mental abilities comparable to those of humans.

The concept of AGI remains somewhat ambiguously defined. There are disagreements among experts regarding what AGI will actually mean in practical terms. Part of the issue with AGI lies in the way it is defined. We face a major challenge, lacking a clear concept and instead encountering a mix of ideas. Nonetheless, this confusion is understandable, as defining something that does not yet exist – and may or may not exist in the future – is inherently difficult.

In the broadest sense, artificial general intelligence refers to AI that has achieved the level of general human intelligence or possesses general capabilities comparable to the average human being. Tom Everitt describes AGI as "a system that surpasses humans in most cognitive tasks" (Everitt, 2018).

The most comprehensive definition of AGI is given by Cameron Hashemi-Pour and Ben Lutkevich, who state (Hashemi-Pour & Lutkevich, 2024): "Artificial General Intelligence (AGI) is a representation of generalized human cognitive abilities through software, such that, when faced with an unfamiliar task, an AGI system could find a solution. The intention of an AGI system is to perform any task that a human being is capable of."

AGI is emphasized as being expected to master human cognitive, that is, non-physical abilities. According to Hashemi-Pour and Lutkevich (Hashemi-Pour & Lutkevich, 2024), an AGI system should be able to demonstrate abilities such as a) abstract thinking; b) background knowledge; c) common sense; d) understanding of cause-and-effect relationships; and e) transfer learning.

Generally, AGI represents the intelligence of a machine capable of performing any human intellectual task as successfully as, or even more successfully than, an average human being. Here, the emphasis is on universality. Unlike ANI, which is highly specialized, AGI is expected to possess a broad range of abilities. It will be autonomous, self-learning, and self-organizing, capable of abstract thinking, quickly learning from specific situations and contexts, solving complex problems, and continuously self-improving and evolving. While AI has not yet reached the AGI level, it is steadily approaching this goal.

Superintelligence is a hypothetical AI concept, one that could emerge after surpassing all capabilities and potential of the average human, becoming an ultra-intelligent AI that is free from all constraints and develops in all directions.

Nick Bostrom defines superintelligence as "an intellect that is much smarter than the best human brains in practically every field, including scientific creativity, general wisdom, and social skills. This definition leaves open the question of how superintelligence is implemented: it could be a digital computer, a network of computers, cultivated cortical tissue, or something else entirely. It also leaves open whether superintelligence is conscious and has subjective experiences" (Bostrom, 2008).

The concept of superintelligence is even more ambiguous than AGI. It is crucial to understand that superintelligence is not just another new technology or a specific kind of technology. It transcends the boundaries and capabilities of technology, defying any attempts at conceptual explanation.

THE RISE OF GENERATIVE AI AND THE POTENTIAL TRANSITION TO AGI

Advancements in NLP technology have brought about a recent breakthrough in large language models (LLMs), leading to a significant leap within the AI field. The rise of generative AI – a type of AI that "can produce various kinds of content, including text, images, audio, and synthetic data – has marked this leap" (Lawton, 2024).

Large language models are part of NLP and fall under the category of generative AI. In the literature, terms like generative AI, LLM, and NLP are not yet clearly delineated.

According to Margaret Rouse, an LLM is "a type of machine learning model that can perform a range of natural language processing tasks, such as text generation and classification, question-answering in a conversational format, and language translation" (Rouse, 2024). An LLM can also be defined as "a type of artificial intelligence algorithm that uses deep learning techniques and massive

datasets to understand, summarize, generate, and predict new content" (Kerner, 2024).

Toloka describes LLM as "a specific application of generative AI". Generative AI is a broader concept within artificial intelligence, encompassing the generation of various types of content.

An important component of large language models, or LLMs, is training – these models are trained on vast datasets to learn and use the data for providing answers and solutions. According to Toloka (Toloka, 2023), "there is no universally recommended figure for how large this training dataset should be, although LLMs can contain a billion parameters or more – parameters in this context are essentially machine learning variables used to train the AI model to draw new conclusions". As LLMs evolve, the training dataset size drastically increases. Training a model involves three stages (Springs, 2024): a) training (engineers pre-train the LLM with large datasets, using information from both open and closed sources); b) fine-tuning; and c) prompt tuning.

A hallmark of AI is its continuous development, evolving toward increasingly sophisticated forms capable of performing more delicate and complex tasks. This evolutionary trajectory moves from the level of narrow, specialized artificial intelligence to the level of artificial general intelligence and, ultimately, superintelligence. Generative AI is a revolutionary step closer to the transition from narrow AI to artificial general intelligence.

In reality, the development after achieving the AGI level is beyond control and serves as an entry into an era of complete uncertainty. The transition to superintelligence and the subsequent entry into the singularity point implies an explosion of intelligence. In 1965, Irving John Good described the concept of an intelligence explosion in relation to the ultra-rapid growth and expansion of AI: "Define an ultra-intelligent machine as a machine that can far surpass all the intellectual activities of any man, however clever. Since the design of machines is one of these intellectual activities, an ultra-intelligent machine could design even better machines; there would then unquestionably be an 'intelligence explosion', and the intelligence of man would be left far behind… Thus the first

ultra-intelligent machine is the last invention that man need ever make, provided that the machine is docile enough to tell us how to keep it under control" (Good, 1965:33).

With today's level of knowledge, it is impossible to predict anything concrete about the development of artificial intelligence beyond the AGI level. The exact criteria for identifying AI as AGI depend on the standards AI must meet to be declared as such. However, the speed at which AGI could develop into the ultimate, most advanced form of AI – superintelligence – may exceed society's ability to recognize, process, and disseminate the news globally.

The emergence of large language models and the development of generative AI led to the launch of the chatbot ChatGPT by OpenAI, which triggered a revolution and an AI race, with other major tech corporations soon releasing competing products. By May 2024, the number of top competitors racing with ChatGPT had grown to over a dozen. Notable applications include Google's Gemini, Microsoft's Copilot, Meta's Meta AI, Apple's OpenELM, Amazon's Amazon Q, IBM's Watson, Claude, and Perplexity AI (Amend, 2024).

The achievement of generative AI, which reaches a peak with chatbots like ChatGPT and others, transforms AI into a creative machine. Although this application of AI currently appears relatively limited by algorithms, its learning capacity is remarkable, and the continuous development and release of increasingly advanced versions are rapidly bringing it closer to the AGI model. With each passing day, we are approaching that moment that will signify the transition of artificial intelligence from narrow AI to AGI. Some authors claim that ChatGPT-4 can be considered an early version of AGI (Bubeck et al.).

Bernard Marr analyzed generative AI concerning its potential transition to AGI (Marr, 2024). Marr compares generative AI to a highly trained parrot capable of mimicking complex patterns and producing various content types without truly understanding the content it creates. For Marr, artificial general intelligence represents a significant leap within the AI field. Thus, AGI will not only be able to create various meaningful content but also understand, innovate, and adapt it as needed. According to Marr (Marr, 2024), the essence of the AGI concept

is "to comprehensively mimic human cognitive abilities, enabling machines to learn and perform a wide range of tasks, from driving cars to diagnosing medical conditions. Unlike any current technology, AGI would not only replicate human actions but also grasp the intricacies and contexts of those actions".

A crucial question is whether generative AI, with its continued super-exponential development, has the capacity to transition into the form of AGI. Can the limitations of generative AI be overcome, and what must be done to facilitate this transition?

The fundamental differences between generative AI and AGI, according to Marr, lie in capabilities, understanding, and application. Regarding capabilities, generative AI, despite impressive results, as Marr highlights, cannot create beyond the boundaries of its programming. In contrast, AGI would be "a powerhouse of innovation capable of understanding and creatively solving problems across various fields, much like a human could" (Marr, 2024). Concerning understanding, generative AI "operates without any real comprehension of its results", generating them based on statistical models and algorithms, while AGI "would need to develop a genuine understanding of the world around it, establishing connections and gaining insights currently beyond the reach of any artificial intelligence" (Marr, 2024). Lastly, in terms of application, AI is widely used in different areas of the economy "to enhance human productivity and stimulate creativity, performing tasks ranging from simple data processing to creating complex content" (Marr, 2024).

ETHICAL IMPLICATIONS AND THE NEED FOR SOCIAL CONTROL OF ARTIFICIAL INTELLIGENCE DEVELOPMENT

In contrast to narrow AI, whose scope is limited to technical tasks, designed to serve specific purposes and perform complex intellectual tasks in place of humans, artificial general intelligence profoundly touches on ethics, interpersonal, and social relationships.

The application of generative AI has irrevocably changed the purpose, objectives, and significance of creation. The emergence of a technology capable

of thinking and producing works, better and more efficiently than humans, fundamentally shifts the traditional perspective on the relationship between technology and society, viewing technology as a tool that assists and benefits humanity.

The market is already witnessing a significant number of books written with AI assistance. One notable example is the novel *1 the Road,* written by Ross Goodwin using artificial intelligence (Hornigold, 2018). The novel was entirely written by AI, with Goodwin merely supervising the process. It was created during a road trip across the USA in March 2017 and published in 2018.

A sensational story emerged from Japan in January 2024 when the young recipient of a prestigious literary award, Rie Kudan, revealed at a press conference that she had used ChatGPT while writing her award-winning novel *Tokyo Tower of Sympathy* (Choi & Annio, 2024). It turned out that AI contributed about 5% of the writing process. This revelation divided the public – some became more interested in the novel, while others protested that it was unfair to other participants who did not use AI.

Many scientists, researchers, and authors have noted that this application is not as reliable as initially believed, observing its behavior when used as a writing tool or to create other types of content. It has become evident that ChatGPT frequently produces incorrect, arbitrary, fabricated, and even fictional, or non-existent results. This phenomenon of generating inaccurate, incorrect, fictional, and fabricated content has been termed "hallucinating". This is particularly problematic in scientific writing. Professor Robin Emsley highlighted serious risks we must consider when using ChatGPT for scientific writing. In describing his experiences using the application to write an introductory article for the journal Schizophrenia, Emsley emphasized that these phenomena (false, fabricated, inaccurate, and unreliable content generated by ChatGPT) should not be called hallucinations because, as he describes, they are "false perceptions" (Emsley, 2023). He concluded, "What I experienced were fabrications and falsifications. The Office of Research Integrity at the U.S. Department of Health and Human Services defines fabrication as making up data or results, and falsification as

manipulating, changing, or omitting data or results so that the research is not accurately represented (https://ori.hhs.gov/definition-research-misconduct). Or, assuming no malintent, confabulations would be a better description, as has been suggested. In any case, the potential consequences are dire. The risk is magnified, first by believing the fabrications and even deceiving established scientists, and second by its tendency to 'double down' when confronted with these inaccuracies. Therefore, use ChatGPT at your own risk. Just as I wouldn't recommend working with a colleague diagnosed with pathological lying (*pseudologia fantastica*), I don't recommend ChatGPT for assistance in scientific writing. While the global push to regulate artificial intelligence is largely driven by the perceived risk of human extinction, it seems to me that the more immediate threat is the infiltration of fictional material into the scientific literature" (Emsley, 2023).

The use of ChatGPT in academia – for writing seminar papers, theses, master's theses, and dissertations – has polarized scientists. Some see it as an efficient tool that simplifies the writing process, while others see it as a threat to authorship integrity and a violation of intellectual property rights. Since the emergence of ChatGPT, numerous studies and analyses have examined its use in scientific writing. Authors Husam Alkaissi and Sami I. McFarlane tested ChatGPT in the field of scientific writing in medicine. They concluded, "While ChatGPT can write credible scientific essays, the data it generates is a mix of truth and complete fabrication. This raises concerns about the integrity and accuracy of using large language models like ChatGPT in academic writing. We propose that policies and practices for evaluating scientific manuscripts for journals and medical conferences should be modified to maintain rigorous scientific standards. We also advocate for the inclusion of AI output detectors in the editing process and the clear disclosure if these technologies are used. The use of large language models in scientific writing remains debatable in terms of ethics and acceptability, as well as the potential to create fake experts in the field of medicine with the potential for harm due to a lack of real expertise and generating expert opinions through ChatGPT" (Alkaissi & McFarlane, 2023).

In all areas of human creativity, AI is displacing people, effortlessly producing even the most complex works. It is penetrating journalism, where editors engage it to write articles, reviews, and critiques. It enters marketing, advertising, and graphic design, generating desired designs, creating new images, or modifying and enhancing existing ones, in prose writing, and even poetry, creating artistic images. Learning and improving, AI relentlessly conquers segments of the vast realm of human creativity.

Due to its boundless capacities for learning, self-correction, and self-improvement, AI radically accelerates and simplifies the creative process. If society does not take control, this will soon lead to a crisis in creativity. With its ability to produce even the most intricate works in much shorter periods, without breaks or errors, strictly adhering to given criteria, and producing creations more beautifully and elegantly than any human can, including the most experienced, qualified and professional individuals, AI may ultimately replace humans in the realm of work and creativity, positioning itself as the superior digital agent.

What will happen to human heritage? How will AI-created art relate to it? Will it build upon and enhance it, or will it completely overshadow and discard it as an archaic and primitive human culture, creating its own unique AI culture?

Who will be considered the author of a work created by AI? This raises questions of authorship and intellectual property rights. Essentially, this brings us to the question of the author's identity. What exactly is the AI that will be the author of a work? To answer this, we need to establish specific rules and criteria for identifying the specific type, class, and form of AI that created the work.

How will issues of falsification, theft, imitation, and copying of authored works be addressed? How will the works created by AI be shared among people and other members of the future society? How will works created by humans and AI collaboratively be regulated?

Somdip Dey analyzed the critical ethical implications that companies engaged in generative AI face and must address. According to Dey (2023), these are:

1. Disinformation and deepfakes;

2. Bias and discrimination;
3. Copyright and intellectual property;
4. Privacy and data security;
5. Accountability.

Dey argues that "the capacity of generative artificial intelligence to produce content that blurs the lines between reality and fiction is alarming". He points to the tremendous damage to reputation that companies can suffer from spreading disinformation, deepfakes, or manipulating information. Dey emphasizes the need for caution and care in training generative AI, ensuring that biased datasets are not used. Additionally, perpetuating or exaggerating social biases "can provoke public anger, legal consequences, and brand damage".

The issue of protecting copyright and intellectual property is also a serious concern that companies dealing with generative AI must consider. Violations of privacy or misuse of personal data pose a significant risk, potentially leading to serious consequences for companies. Generative AI presents a considerable risk to citizens' privacy by enabling unauthorized use of personal data, generating synthetic profiles identical to original ones, and sharing and using private information without the knowledge or consent of individuals.

Finally, it is essential to establish accountability for spreading false news, disinformation, hate speech, manipulation, and other forms of abuse associated with generative AI. Dey suggests that solid guidelines and policies of conduct be established, clearly outlining what is permissible and what is not, and insisting on accountability at all stages of activity.

In a survey conducted by the Pew Research Center between July 3 and August 5, 2019, leading experts and researchers in new technologies shared their views on the future impact of digital technology on democracy and social innovation (Pew Research Center, 2020).

The results indicated that many experts are concerned about the future impact of technology, believing that technology often creates more problems than it solves. A significant number of experts cited disinformation and fake news as serious issues in digital spaces. Many recognized the growing need

for privacy protections, while some highlighted rising problems related to community fragmentation and the distancing effects of technology, emphasizing the need for more organic, personal, face-to-face interactions. Finally, many experts see a challenging and complex road ahead but hold hope for the future.

These striking findings vividly confirm what remains less discussed or overlooked – the social aspect of AI. What are the social consequences of AI's rapid progress? How will the increasingly widespread application of AI affect human behavior, habits, lifestyle, mental health, and communication? What will the impacts be on social life?

Fortunately, there are growing calls for re-evaluating the trend of AI enthusiasm and examining the other side of AI expansion. This leads us to what is known as the AI Dilemma. The AI Dilemma emphasizes the importance of recognizing the gap between the technical and social aspects of AI development. It underscores the urgency of ethical constraints on AI and the necessity of social control over AI to ensure safe, responsible, and fair development that benefits society as a whole.

Tristan Harris and Aza Raskin, AI researchers, discussed the AI Dilemma during a presentation in San Francisco on March 9, 2023, explaining that existing AI capabilities already pose a massive threat to society as a whole (Center for Humane Technology, 2023).

They mentioned that humanity has had two major encounters with AI. The first encounter with AI was through social media. The second encounter is now taking place with generative AI, namely chatbots. The two experts warned that while people initially touted the many benefits of social media, it simultaneously led to a series of severe social issues. Harris stated: "So now, we are literally in contact with AI every day – a very simple technology that simply calculates which photo, which video, which cat video, or which birthday post to show your nervous system to keep you scrolling. But that pretty simple technology was enough, in our first contact with AI, to bring humanity information overload, addiction, news tracking obsession, child sexualization, shortened attention spans, polarization, fake news, and the breakdown of democracy. And no one

planned for these things to happen. It was just a bunch of engineers saying they were trying to maximize engagement. It seemed so harmless. And so, in our first contact with social media, humanity lost" (Center for Humane Technology, 2023).

The authors of the AI Dilemma presentation emphasized how, during both encounters, we were drawn to the benefits and capabilities of AI systems, unaware of the hidden negative effects. In other words, we saw only one side and were oblivious to the existence of these other consequences, which were hidden. Now, in this second encounter, the negative effects could be devastating. A significant threat from this second encounter with AI is the invasion of privacy and intimacy, implying a range of frightening possibilities, from superior persuasion capabilities and reading our thoughts and feelings to complete control over our minds.

Regulating AI-related policies and legally controlling AI is one of the most serious and complex issues. In recent years, society, led by corporate executives developing AI, prominent experts, researchers, scientists, writers, and other stakeholders, has taken numerous significant steps to raise public awareness of the risks posed by advanced AI and to make decisions, initiatives, and measures for monitoring, analyzing, and socially controlling the use of AI.

The Center for AI Safety, whose mission is to reduce AI-related societal risks, has published a classification of catastrophic risks that AI poses to society, divided into four categories:

• Malicious Use: Risks related to the malicious use of AI by individuals to cause large-scale harm. This includes bioterrorism (using AI to create new pandemics and discover biological or chemical weapons), AI for propaganda, manipulation, censorship, and surveillance, among others.

• AI Race: This race has two forms: a) AI for military purposes (arms race) and b) corporate competition. The AI race for military advantage could lead to unplanned conflict or war. Autonomous weapons and cyber warfare could spiral out of control, resulting in catastrophic consequences. In the corporate context, companies, driven by competition, might face challenges to automating

human labor, leading to mass unemployment and dependency on AI systems. With enabling autonomous replication, AI may evolve into higher forms that will be increasingly challenging to control.

• Organizational Risks: Risks associated with organizations arise mainly from carelessness, irresponsible business practices, and insufficient oversight. It is essential that organizations establish a culture that supports responsibility and safety in AI applications. Nothing should be left to chance. Key information crucial for AI development might leak, or organizational oversights might allow AI to behave unpredictably.

• Deceptive AI: These risks emerge when AI becomes more capable, potentially breaking free from control. Such AI could manipulate information and deceive, resist shutdown, make autonomous decisions on objectives, or deviate from initial goals.

Awareness of the risks associated with unchecked AI growth, which could transform into an autonomous force making decisions contrary to ethical principles, has prompted the EU to adopt policies and regulations aimed at developing reliable and acceptable AI that benefits, rather than harms, human and economic and social progress. The EU's approach to AI is based on understanding AI's operational risks, with a human-centered perspective, focusing on excellence and trustworthy AI.

The EU has formulated a European approach to AI. The goal of the European AI approach is to enable AI to be maximally effective and applied across a broad spectrum of economic and social fields, with a focus on two areas: a) AI excellence; and b) trusted AI.

The most significant breakthrough in AI policy and regulation occurred on April 21, 2021, when the EU proposed the Artificial Intelligence Act, formally known as "Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts".

The AI Act was enthusiastically received in the EU, and passed on April 19, 2024, marking it as the first global legislation of its kind (European Parliament,

2024). This legislation pays close attention to the ethical challenges posed by AI, as well as the opportunities for AI limitations and control to ensure that AI development serves humanity and society as a whole.

At the core of the AI Act is a risk-based approach, which categorizes AI systems into four levels of risk, presented as a pyramid based on risk height: 1) minimal risk level; 2) limited risk level; 3) high-risk level; and 4) unacceptable risk level (Holland & Knight, 2024).

The categorization of AI risk levels considers two factors: a) the sensitivity of the data involved; and b) the specific use case or AI application. The law explicitly bans AI applications that fall into the "unacceptable risk" category. These prohibited applications include marketing that: a) involves AI systems using manipulative, deceptive, and/or subliminal techniques to influence a person to make a decision they would not otherwise make, causing harm to themselves or others; b) exploits a person's vulnerability due to age, disability, or specific socio-economic situation to influence their behavior, potentially causing significant harm to themselves or others; c) uses biometric data to categorize individuals based on their race, political opinions, union membership, religious or philosophical beliefs, sexual life, or sexual orientation; and d) creates or expands facial recognition databases through indiscriminate facial image capture from the internet or CCTV footage.

It is important to note that the AI Act is not an isolated legal and policy instrument but is supported by other significant measures and policies. This package of measures, instruments, and documents includes the AI Innovation Package and the Coordinated Plan on AI. The Coordinated Plan on AI aims to accelerate AI investment to drive the recovery of the EU economy, promote the full implementation of AI strategies and programs within the EU, and coordinate AI policy to tackle global challenges.

The AI Act is designed to ensure AI control and mitigate associated risks. Its goal is to build trust in AI, reflecting a European approach that prioritizes people. Furthermore, this approach to AI aims to position the EU as a global competitor in the AI field.

Numerous documents show that the EU strives to adhere firmly to fundamental European values and principles, emphasizing a human-centric approach to technologies, safeguarding people's integrity and rights, sustainability, and technology access that benefits humanity for the general good and prosperity. In this way, the EU becomes a global leader and model in AI governance and regulation, demonstrating to other economic and social entities how to address urgent ethical, safe, and responsible AI use and ensure harmonious and prosperous economic, technological, and social development in the future.

CONCLUSION

This paper has explored the ethical implications of artificial intelligence development in terms of the transition to artificial general intelligence (AGI). The rise of generative AI has further intensified AI's growth and development, spurred a global AI race, and confronted society with some critical issues. We find ourselves at a crossroads, facing a societal choice: whether to pursue profit, wealth, and the power of the super-wealthy individuals who control AI or choose to limit and control AI in line with an ethical approach to AI use, ensuring that AI's growth and development benefit humanity. This is a momentous question and an opportunity to keep technology and development under control so that society as a whole can benefit from it.

Today, as a society, we are on the brink of an AI crisis, facing the AI Dilemma. Due to the unchecked growth and uncontrolled infusion of AI into various sectors of the economy and society, we risk entering a super-exponential spiral of AI growth, which could soon lead to AGI and AI autonomy, with massive societal consequences. Even at this early stage of generative AI development, many weaknesses and challenges have been identified. It is crucial to reconsider AI's future development toward achieving AGI and superintelligence. If we allow this development to progress unchecked, we will enter an era of permanent uncertainty, with unimaginable consequences and risks for people and society as a whole.

Therefore, it is essential to involve all economic and social stakeholders and take ethical, legal, and political regulatory measures at both the state and global levels to ensure the safe, socially responsible, and beneficial use of AI that does not threaten privacy, fundamental human rights, and democracy.

LITERATURE

1. Alkaissi, Husam, and Sami I. McFarlane. 2023. "Artificial Hallucinations in ChatGPT: Implications in Scientific Writing." *Cureus* 15 (2): e35179. https://doi.org/10.7759/cureus.35179.
2. Amend, Robert. 2024. "9 ChatGPT Competitors: Who Will Win the AI Race?" *24/7 Wall St.*, May 1. https://247wallst.com/apps-software/2024/05/01/9-chatgpt-competitors-who-will-win-the-ai-race/.
3. Bostrom, Nick. 2003. "Ethical Issues in Advanced Artificial Intelligence." https://nickbostrom.com/ethics/ai.
4. Bostrom, Nick. 2008. "How Long Before Superintelligence?" https://nickbostrom.com/superintelligence.
5. Bubeck, Sébastien, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Paul Lee, et al. 2023. "Sparks of Artificial General Intelligence: Early Experiments with GPT-4." *arXiv*. https://arxiv.org/pdf/2303.12712.
6. Center for Humane Technology. 2023. "The A.I. Dilemma." *YouTube*, March 9. https://www.youtube.com/watch?v=xoVJKj8lcNQ.
7. Center for AI Safety. 2023. "Risks from AI: An Overview of Catastrophic AI Risks." https://www.safe.ai/ai-risk#malicious-use.
8. Choi, Connie, and Francesca Annio. 2024. "The Winner of a Prestigious Japanese Literary Award Has Confirmed AI Helped Write Her Book." *CNN Style*, January 19. https://edition.cnn.com/2024/01/19/style/rie-kudan-akutagawa-prize-chatgpt/index.html.
9. Dey, Somdip. 2023. "Which Ethical Implications of Generative AI Should Companies Focus On?" *Forbes*, October 17. https://www.forbes.

com/sites/forbestechcouncil/2023/10/17/which-ethical-implications-of-generative-ai-should-companies-focus-on/.

10. European Parliament. 2024. *Artificial Intelligence Act*. https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138-FNL-COR01_EN.pdf.
11. Emsley, Robin. 2023. "ChatGPT: These Are Not Hallucinations – They're Fabrications and Falsifications." *Schizophrenia* 9: 52. https://doi.org/10.1038/s41537-023-00379-4.
12. Everett, Tom. 2018. *Towards Safe Artificial General Intelligence*. Doctoral thesis, Stanford University. https://www.tomeveritt.se/papers/2018-thesis.pdf.
13. Good, I. J. 1965. "Speculations Concerning the First Ultraintelligent Machine." *Advances in Computers* 6.
14. Hashemi-Pour, Cameron, and Ben Lutkevich. 2024. "What Is Artificial General Intelligence (AGI)?" *TechTarget*, May. https://www.techtarget.com/searchenterpriseai/definition/artificial-general-intelligence-AGI.
15. Holland & Knight IP/Decode Blog. 2024. "The European Union's AI Act: What You Need to Know." March 15. https://www.hklaw.com/en/insights/publications/2024/03/the-european-unions-ai-act-what-you-need-to-know.
16. Hornigold, Tom. 2018. "The First Novel Written by AI Is Here—and It's as Weird as You'd Expect It to Be." *Singularity Hub*, October 25. https://singularityhub.com/2018/10/25/ai-wrote-a-road-trip-novel-is-it-a-good-read/.
17. Investopedia. 2020. "Artificial Intelligence (AI)." Accessed June 14, 2024. https://www.investopedia.com/terms/a/artificial-intelligence-ai.asp.
18. Kerner, Sean Michael. 2024. "What Are Large Language Models (LLMs)?" *TechTarget*. Accessed June 13, 2024. https://www.techtarget.com/whatis/definition/large-language-model-LLM.

19. Lawton, Graham. 2024. "What Is Generative AI? Everything You Need to Know." *TechTarget*. https://www.techtarget.com/searchenterpriseai/definition/generative-AI.
20. Marr, Bernard. 2024. "The Important Difference Between Generative AI and AGI." *Forbes*, May 8. https://www.forbes.com/sites/bernardmarr/2024/05/08/the-important-difference-between-generative-ai-and-agi/.
21. Matthews, Dylan. 2024. "How AI Could Explode the Economy." *Vox*, May 26. https://www.vox.com/future-perfect/24108787/ai-economic-growth-explosive-automation.
22. McCarthy, John. 2007. "What Is Artificial Intelligence?" *Computer Science Department, Stanford University*. Accessed June 6, 2024. http://jmc.stanford.edu/articles/whatisai/whatisai.pdf.
23. McKinsey & Company. 2024. "What Is AI (Artificial Intelligence)?" https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-ai.
24. Pew Research Center. 2020. "Experts Predict More Digital Innovation by 2030 Aimed at Enhancing Democracy." June 30. https://www.pewresearch.org/internet/2020/06/30/tech-causes-more-problems-than-it-solves/.
25. Rouse, Margaret. 2024. "Large Language Model (LLM)." *Techopedia*, April 25. https://www.techopedia.com/definition/34948/large-language-model-llm.
26. Roser, Max. 2022. "The Brief History of Artificial Intelligence: The World Has Changed Fast — What Might Be Next?" *OurWorldInData.org*. https://ourworldindata.org/brief-history-of-ai.
27. Takyar, Ankit. 2024. "From Data to Decisions: A Guide to the Core AI Technologies." *LeewayHertz*. https://www.leewayhertz.com/key-ai-technologies/.
28. Toloka. 2023. "Difference Between AI, ML, LLM, and Generative AI." https://toloka.ai/blog/difference-between-ai-ml-llm-and-generative-ai/.
29. World Economic Forum. 2017. *Global Risks Report 2017*. http://www3.weforum.org/docs/GRR17_Report_web.pdf.